

Klasifikasi perilaku konsumen dalam penggunaan *E-commerce* menggunakan algoritma KNN

¹Saut Dohot Siregar, ²Ghaniya Syifa Talitha, ³Louis Alonso, ⁴Navita Puan chairy, ⁵Jannes Pakpahan, ⁶Richie Anderson Joe

^{1,2,3,4,5,6} Universitas Prima Indonesia

¹sautdohotsiregar@gmail.com, ²ghaniyagst@gmail.com, ³louisalonso114@gmail.com,
⁴navitapuan6@gmail.com, ⁵pakpahanjannes0@gmail.com, ⁶Richie.a.joe@gmail.com

Abstrak. Penelitian ini bertujuan untuk menganalisis perilaku konsumen dalam penggunaan e-commerce di Indonesia dengan memanfaatkan algoritma K-Nearest Neighbor (KNN). Dalam penelitian ini, data yang digunakan meliputi beberapa variabel penting, seperti frekuensi belanja, jenis produk yang dibeli, dan metode pembayaran yang dipilih oleh konsumen. Algoritma KNN diterapkan untuk mengidentifikasi pola dan mengklasifikasikan perilaku konsumen berdasarkan data yang ada. Hasil analisis menunjukkan bahwa algoritma KNN mampu memberikan klasifikasi yang akurat mengenai perilaku konsumen, serta mampu mengelompokkan konsumen berdasarkan karakteristik tertentu, seperti kebiasaan berbelanja dan preferensi produk. Temuan ini memberikan wawasan yang berguna bagi perusahaan e-commerce dalam merancang strategi pemasaran yang lebih terarah dan personal, dengan memanfaatkan pemahaman yang lebih mendalam tentang pola konsumen. Selain itu, penelitian ini juga menyoroti potensi penggunaan algoritma pembelajaran mesin, seperti KNN, untuk menganalisis dan memahami data konsumen dalam konteks e-commerce. Dengan demikian, penelitian ini berkontribusi pada pengembangan pemahaman mengenai perilaku konsumen dalam e-commerce, serta pentingnya penerapan teknik analisis data canggih untuk meningkatkan efektivitas strategi pemasaran dan pengalaman konsumen.

Kata Kunci: K-Nearest Neighbor, Klasifikasi, Perilaku Konsumen, e-commerce

1. Pengantar

Kemajuan teknologi informasi dan komunikasi yang berkembang telah membawa perubahan yang signifikan dalam berbagai bidang kehidupan manusia, salah satunya terkait dengan cara kita berbelanja. *e-commerce*, juga dikenal sebagai perdagangan elektronik, sudah menjadi sebuah tren global yang terus berkembang dan semakin diminati oleh masyarakat. berdasarkan laporan Statista, nilai transaksi dari *e-commerce* global telah mencapai \$ 1 triliun per tahun dan diperkirakan akan terus naik seiring dengan meningkatnya jumlah pengguna internet dan kepercayaan konsumen terhadap belanja online.

E-commerce juga mengalami peningkatan yang cepat di Indonesia. menurut Laporan dari Asosiasi *e-commerce* Indonesia (idEA) menunjukkan bahwa pertumbuhan dalam jumlah transaksi *e-commerce* di Indonesia dari tahun ke tahun terus meningkat. Namun, terlepas dari semakin populernya

e-commerce di kalangan masyarakat, ada beberapa faktor yang memengaruhi perilaku konsumen saat menggunakan *e-commerce*.

Sangat penting bagi perusahaan bisnis untuk memiliki pemahaman yang kuat tentang perilaku konsumen dalam penggunaan *e-commerce* untuk mengembangkan strategi pemasaran yang efektif, meningkatkan pengalaman pelanggan dan menumbuhkan loyalitas pelanggan.

Memanfaatkan algoritma pembelajaran mesin (machine learning) untuk menganalisis data merupakan pendekatan yang dapat digunakan untuk memperoleh wawasan tentang perilaku konsumen. Algoritma K-Nearest Neighbor (KNN) merupakan algoritma yang umumnya digunakan dalam analisis data karena kemampuannya mengklasifikasikan data berdasarkan kedekatan atau kesamaannya dengan data lain. Pemanfaatan algoritma KNN membantu dalam mengklarifikasi dan menganalisis perilaku konsumen dalam penggunaan *e-commerce* berdasarkan berbagai variabel seperti frekuensi pembelian, jenis barang yang dibeli, metode pembayaran yang digunakan dan banyak lagi.

2. Tinjauan Literatur

Penelitian tentang perilaku konsumen dalam penggunaan e-commerce telah banyak dilakukan oleh para peneliti[9]. Beberapa studi sebelumnya telah mengidentifikasi berbagai faktor yang mempengaruhi perilaku konsumen dalam berbelanja online, seperti demografi (Alreck & Settle, 2002; Korgaonkar & Wolin, 1999), motivasi (Childers et al., 2001; Rohm & Swaminathan, 2004), pengalaman sebelumnya (Monsuwé et al., 2004), dan preferensi (Donthu & Garcia, 1999; Swinyard & Smith, 2003)[10].

Beberapa penelitian telah menerapkan teknik klasifikasi untuk menganalisis perilaku konsumen dalam e-commerce[11]. Misalnya, Reutskaja dan Hogarth (2009) menggunakan model logit untuk mengklasifikasikan konsumen berdasarkan preferensi mereka terhadap atribut e-commerce. Sementara itu, Huang dan Benyoucef (2013) menerapkan analisis cluster untuk mengidentifikasi segmen konsumen berdasarkan perilaku mereka dalam komunitas online[11].

Algoritma KNN merupakan salah satu metode klasifikasi yang telah banyak digunakan dalam penelitian perilaku konsumen[12]. Beberapa studi sebelumnya telah menerapkan KNN untuk mengklasifikasikan konsumen berdasarkan karakteristik demografis (Ting et al., 2013), pola pembelian (Vellido et al., 1999), dan preferensi terhadap produk (Sarkar & Chakrabarti, 2011)[13].

Meskipun beberapa penelitian telah menerapkan algoritma KNN untuk menganalisis perilaku konsumen, masih sedikit studi yang secara spesifik mengkaji penggunaan KNN dalam mengklasifikasikan perilaku konsumen dalam konteks e-commerce[14]. Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengeksplorasi penerapan algoritma KNN untuk mengklasifikasikan perilaku konsumen dalam penggunaan e-commerce berdasarkan berbagai variabel yang mempengaruhi[15]

3. Metode penelitian

1. Pengumpulan Data:

- a. Identifikasi dan kumpulkan data perilaku konsumen dalam penggunaan e-commerce.
- b. Data bisa didapat dari survei, observasi, atau sumber data sekunder yang berkaitan.

2. Persiapan Data:

- a. Lakukan preprocessing data, seperti missing values, outliers, dan normalisasi data.
- b. Bagi data menjadi set pelatihan dan set pengujian

3. Pemodelan dengan Algoritma K-Nearest Neighbors (KNN):

- a. Tentukan jumlah tetangga yang terdekat (k) yang akan diterapkan dalam algoritma KNN.
- b. Implementasikan algoritma KNN untuk mengklasifikasikan perilaku konsumen berdasarkan bermacam-macam fitur yang ada.

4. Evaluasi Kinerja Model:

- a. Hitung metrik evaluasi, seperti akurasi, presisi, recall, dan F1-score, sebagai cara untuk mengukur atau menilai kinerja model KNN dalam mengklasifikasikan perilaku pengguna atau konsumen.

- b. Lakukan validasi silang (cross-validation) untuk memperkirakan kinerja model lebih akurat.

5. Analisis Hasil:

- a. Interpretasikan hasil klasifikasi dan analisis pola perilaku konsumen yang teridentifikasi.
- b. Identifikasi beberapa faktor yang mempengaruhi tindakan konsumen dalam menggunakan e-commerce.

6. Dokumentasi dan Pelaporan:

- a. Dokumentasikan semua langkah-langkah penelitian, hasil, dan temuan dalam bentuk laporan atau artikel ilmiah.
- b. Tampilkan hasil penelitian dalam bentuk tabel, grafik, atau visualisasi yang jelas dan informatif.

4. Hasil Penelitian

1. Pengumpulan Data

Data ini meliputi faktor-faktor pengaruh perilaku pengunjung online dan niat membeli. Data ini berisikan durasi kunjungan, nilai halaman, bounce rate, keluar rate, special day, bulan, sistem operasi, browser, wilayah, type traffic, type visitor, weekend dan membayar. Untuk memulai analisis, kita membaca dan menampilkan lima baris pertama data untuk pemahaman struktur dan isi data.

```
In [5]: online = pd.read_csv('online_shoppers_intention.csv')
        online.head()
```

```
Out[5]:
```

Related_Duration	BounceRates	ExitRates	PageValues	SpecialDay	Month	OperatingSystems	Browser	Region	TrafficType	VisitorType	Weekend	Revenue
0.000000	0.20	0.20	0.0	0.0	Feb	1	1	1	1	Returning_Visitor	False	False
84.000000	0.00	0.10	0.0	0.0	Feb	2	2	1	2	Returning_Visitor	False	False
0.000000	0.20	0.20	0.0	0.0	Feb	4	1	0	3	Returning_Visitor	False	False
2.696667	0.05	0.14	0.0	0.0	Feb	3	2	2	4	Returning_Visitor	False	False
827.500000	0.02	0.06	0.0	0.0	Feb	3	3	1	4	Returning_Visitor	True	False

2. Persiapan Data

Sebelum melakukan analisis lebih lanjut kita harus menyiapkan data yang telah dikumpulkan. Persiapan data mencakup pengecekan kualitas data, menangani data yang hilang, dan mentransformasi data sehingga siap dianalisis. Dataset dalam penelitian ini adalah online_shoppers_intention.csv, berisi data pengunjung toko beserta semua faktor pengaruh niat membeli atau tidak dan digunakan untuk penelitian.

```
In [8]: # Import libraries
import pandas as pd
from sklearn.preprocessing import LabelEncoder, StandardScaler
from sklearn.model_selection import train_test_split

# Load dataset
df = pd.read_csv('online_shoppers_intention.csv')

# Exploratory Data Analysis
print(df.info())
print(df.describe())
print(df.isnull().sum())

# Handle missing values
df = df.dropna()

# Encode categorical features
le = LabelEncoder()
df['Month'] = le.fit_transform(df['Month'])
df['VisitorType'] = le.fit_transform(df['VisitorType'])
df['Weekend'] = le.fit_transform(df['Weekend'])

# Scale numerical features
scaler = StandardScaler()
df[['Administrative', 'Informational', 'ProductRelated', 'BounceRates', 'ExitRates', 'PageValues', 'SpecialDay']] = scaler.fit_transform(df[['Administrative', 'Informational', 'ProductRelated', 'BounceRates', 'ExitRates', 'PageValues', 'SpecialDay']])

# Split data into features and target
X = df.drop('Revenue', axis=1)
y = df['Revenue']
```

	Administrative	Administrative_Duration	Informational
count	12330.000000	12330.000000	12330.000000
mean	2.335168	88.818613	8.583589
std	3.322784	176.779107	1.270156
min	0.000000	0.000000	0.000000
25%	0.000000	0.000000	0.000000
50%	1.000000	3.000000	0.000000
75%	4.000000	93.246258	0.000000
max	27.000000	3398.750000	24.000000

	Informational_Duration	ProductRelated	ProductRelated_Duration
count	12330.000000	12330.000000	12330.000000
mean	81.072398	31.731488	1194.708230
std	140.749204	44.475503	1913.609280
min	0.000000	0.000000	0.000000
25%	0.000000	7.000000	104.137500
50%	0.000000	18.000000	508.036905
75%	0.000000	38.000000	1484.197214
max	2549.375000	705.000000	83973.522238

	BounceRates	ExitRates	PageValues	SpecialDay
count	12330.000000	12330.000000	12330.000000	12330.000000
mean	0.822291	0.861073	9.882258	0.001427
std	0.040460	0.048597	10.268437	0.108927
min	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.014204	0.000000	0.000000
50%	0.001112	0.025150	0.000000	0.000000
75%	0.818813	0.898088	0.000000	0.000000

3. Pemodelan dengan Algoritma K-Nearest Neighbors (KNN)

Model yang memprediksi perilaku atau hasil tertentu bisa dibuat dengan data mentah yang sudah disiapkan untuk analisis. Prediksi dalam penelitian ini menggunakan algoritma K-Nearest Neighbors (KNN), salah satu algoritma [i]machine learning[/i] yang paling sederhana dan paling umum dipakai untuk klasifikasi dan regresi. Algoritma ini membuat titik data prediksinya dengan menggunakan informasi dari titik-titik data terdekat (tetangga-tetangga) untuk menetapkan klasifikasi atau nilai titik data prediksinya.

a. Data dipisahkan menjadi training set dan testing set

Training Set:						
	Administrative	Informational	ProductRelated	BounceRates	ExitRates	
7923	0	0	47	0	0	
5814	0	0	3	0	0	
6826	1	0	11	0	0	
6021	0	0	5	0	0	
6267	0	0	4	0	0	

	PageValues	SpecialDay	Month	OperatingSystem	Browser	Region
7923	0	0	0	1	3	0
5814	0	0	3	0	0	0
6826	0	0	3	1	1	5
6021	0	0	8	0	0	2
6267	0	0	0	2	1	3

	TrafficType	VisitorType	Weekend
7923	1	0	0
5814	1	2	1
6826	12	2	0
6021	3	2	1
6267	3	0	0

Testing Set:					
	Administrative	Informational	ProductRelated	BounceRates	ExitRates
6381	2	0	7	0	0
6333	0	0	1	0	0
6314	0	0	7	0	0
6828	1	1	7	0	0
6673	4	0	10	0	0

b. model dilatih menggunakan training set, kemudian model dites dengan testing set, lalu pengevaluasian model KKN dilakukan dengan rumus-rumus statistika.

```

Accuracy: 0.884822708848227
Classification Report:
              precision    recall  f1-score   support

 False       0.91      0.96      0.93      2855
  True       0.72      0.50      0.59       411

 accuracy    0.88      0.88      0.88      2466
 macro avg   0.81      0.73      0.76      2466
 weighted avg 0.87      0.88      0.88      2466

Confusion Matrix:
[[1973  82]
 [ 204 207]]

```

4. Evaluasi kinerja model

Setelah model jadi, kinerja model perlu dievaluasi. Dalam evaluasi ini, kita melihat seberapa baik model dapat memprediksi tingkah laku pembeli berdasarkan data yang tidak pernah diberikan untuk latihan. Dalam penelitian ini, gambaran menyeluruh tentang kinerja model didapat menggunakan berbagai metrik evaluasi.

a. Akurasi

- Akurasi model KNN adalah 0,88

b. Laporan klasifikasi

- Laporan klasifikasi untuk model KNN

```

Classification Report:
              precision    recall  f1-score   support

 False       0.91      0.96      0.93      2855
  True       0.72      0.50      0.59       411

 accuracy    0.88      0.88      0.88      2466
 macro avg   0.81      0.73      0.76      2466
 weighted avg 0.87      0.88      0.88      2466

```

c. Matriks konfusi

- untuk model KNN

```

Confusion Matrix:
[[1973  82]
 [ 204 207]]

```

5. Analisis hasil

a. Data Preprocessing

- Pemuatan Dataset: Dataset "Online_shoppers_intention.csv" dimuat menggunakan Pandas.
- Pendefinisian Variabel Target dan Fitur: Variabel target merupakan "Revenue", dan fitur-fitur yang digunakan merupakan "Administrative", "Informational", "ProductRelated", "BounceRates", "ExitRates", "PageValues", "SpecialDay", "Month", "OperatingSystems", "Browser", "Region", "TrafficType", "VisitorType", dan "Weekend".
- Mengubah Variabel Kategorikal ke Numerik: Variabel kategorikal diubah menjadi numerik menggunakan LabelEncoder dari Sklearn.
- Menghapus Baris dengan Nilai Tidak Ada: Baris dengan nilai tidak ada dihapus menggunakan fungsi dropna dari Pandas.
- Menghapus Baris dengan Nilai Tidak Konsisten: Baris dengan nilai tidak konsisten (misalnya, nilai negatif) dihapus menggunakan kondisi logika.
- Menghapus Baris dengan Nilai Ekstrem: Baris dengan nilai Ekstrem ([i]outlier[i]) dihapus menggunakan kondisi logika.
- Menghapus Baris dengan Nilai Duplikat: Baris dengan nilai duplikat dihapus menggunakan fungsi drop_duplicates dari Pandas.

b. Pembagian Data

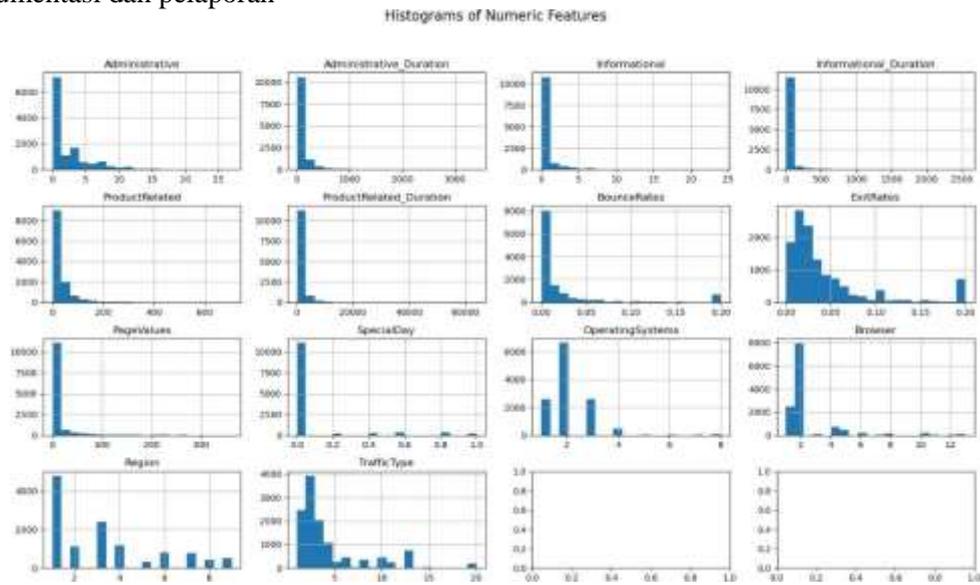
- Membagi Data ke Training dan Testing: Data dibagi menjadi bagian pelatihan (80%) dan uji (20%) menggunakan fungsi train_test_split dari Sklearn.

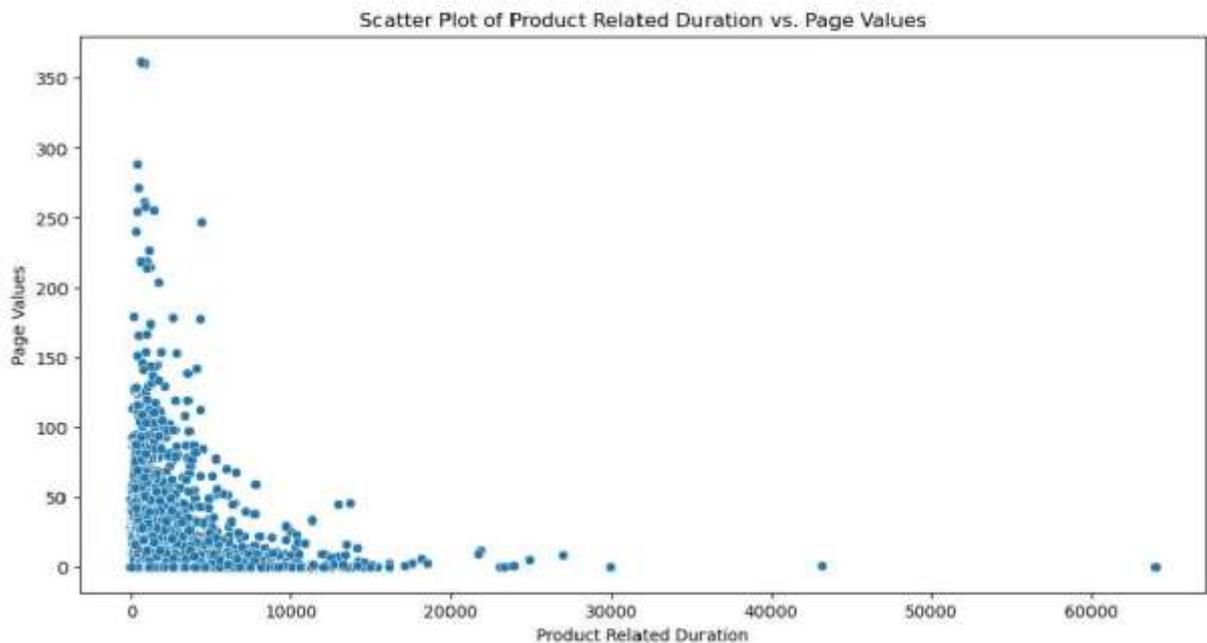
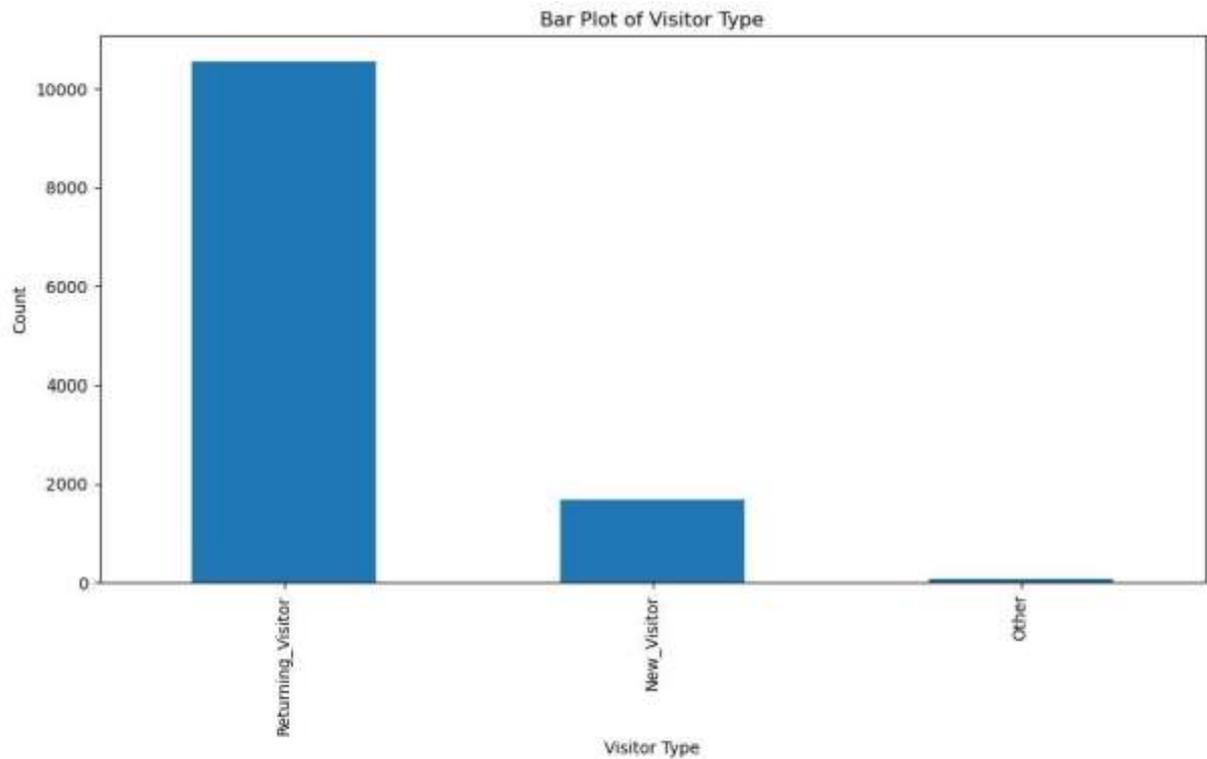
c. Skalasi Data

- Skalasi Data: Data diskalasi menggunakan StandardScaler dari Sklearn.

- d. Model K-Nearest Neighbor (KNN)
 - Menginisialisasi dan Melatih Model KNN: Model KNN diinisialisasi dan dilatih menggunakan data pelatihan.
- e. Evaluasi Model
 - Mengevaluasi Model: Model dievaluasi menggunakan fungsi `accuracy_score`, `classification_report`, dan `confusion_matrix` dari Sklearn.
- f. Tuning Hyperparameter
 - Tuning Hyperparameter: Hyperparameter model KNN dituning menggunakan `GridSearchCV` dari Sklearn.

6. Dokumentasi dan pelaporan





5. Pembahasan

Penelitian ini membahas tentang bagaimana perilaku seorang konsumen dalam kebijakannya menggunakan E-commerce di Indonesia serta bagaimana peran yang dilakukan K-Nearest Neighbor sebagai algoritma pembelajaran mesin. Oleh karena itu, analisis dari algoritma tersebut efektif, karena dapat memungkinkan perusahaan berbasis E-commerce yang beroperasi wilayah Indonesia dapat mengklasifikasikan konsumen dalam parameter frekuensi belanja, jenis produk yang dibeli, dan metode pembayarannya. Adapun riset ini memiliki implikasi signifikan bagi sebuah bisnis, karena peran K-Nearest Neighbor dalam memberi dukungan terhadap strategi pemasaran E-commerce yang

efektif. Juga, penugasan ini memiliki beberapa kontribusi yang berguna dalam pemahaman terhadap perilaku seorang konsumen dan bagaimana algoritma pembelajaran mesin membantu dalam menganalisis data.

Seperti yang diidentifikasi penelitian lainnya, dengan temuan yang sebelumnya dihasilkan atas faktor-faktor yang memperdayai perilaku seorang konsumen di dalam belanja online mencoba untuk mengkhususkan elemen tersebut dan melengkapi bukti dengan pengaplikasian algoritma K-Nearest Neighbor, yang dapat melihat perilaku seorang konsumen menggunakan E-commerce.

Daftar Pustaka

- [1] I. S. Milal, M. H. M. Hasanudin, M. A. Nur Azhari, R. A. Nugraha, N. Agustina, and S. E. Damayanti, "Klasifikasi Teks Review Pada E-Commerce Tokopedia Menggunakan Algoritma Svm," *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 5, no. 1, pp. 34–45, 2023, doi: 10.53580/naratif.v5i1.191.
- [2] M. Afdal and L. R. Elita, "Penerapan Text Mining Pada Aplikasi Tokopedia Menggunakan Algoritma K- Nearest Neighbor," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 8, no. 1, p. 78, 2022, doi: 10.24014/rmsi.v8i1.16595.
- [3] M. Kondoj, H. Langi, Y. Putung, and A. Kumaat, "Model E-Commerce Untuk Meningkatkan Daya Saing UMKM Dalam Ekosistem Kewirausahaan Digital di Sulawesi Utara," *Technomedia J.*, vol. 8, no. 2, pp. 221–234, 2023, doi: 10.33050/tmj.v8i2.2089.
- [4] K. R. Putra and I. F. Rahman, "Pemanfaatan metode collaborative filtering dengan algoritma KNN pada sistem rekomendasi produk," *MIND (Multimedia Artif. Intell. Netw. Database) J.*, vol. 9, no. 1, pp. 113– 123, 2024.
- [5] W. Prianata, "Dampak Pengambilan Sampel Data untuk Optimalisasi Data Tidak Seimbang pada Klasifikasi Penipuan Transaksi E-Commerce Wowon Priatna," *Indones. J. Comput. Sci. Attrib.*, vol. 13, no. 2, p. 3070, 2024.
- [6] K. M. Hindrayani, A. Anjani, and A. L. Nurlaili, "Penerapan Machine Learning pada Penjualan Produk UMKM : Studi Literatur," *Semin. Nas. Sains Data*, vol. 2021, no. 2021, pp. 19–23, 2021.
- [7] M. Saifurridho, M. Martanto, and U. Hayati, "Analisis Algoritma K-Nearest Neighbor terhadap Sentimen Pengguna Aplikasi Shopee," *J. Inform. Terpadu*, vol. 10, no. 1, pp. 21–26, 2024, doi: 10.54914/jit.v10i1.1054.
- [8] A. Wandani, "Sentimen Analisis Pengguna Twitter pada Event Flash Sale Menggunakan Algoritma K- NN, Random Forest, dan Naive Bayes," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 651–665, 2021.
- [9] S. Pramuwidya, R. Y. Fa'rifah, and O. N. Pratiwi, "Klasifikasi Review Customer E-Commerce Menggunakan Metode K-Nearest Neighbor (Studi Kasus : Bukalapak)," *e-Proceeding Eng.*, vol. 10, no. 3, pp. 3207– 3212, 2023.
- [10] N. Saurina, "Analisis Sentimen Ulasan Pelanggan Batik Ecoprint Menggunakan Naive Bayes Dan Knn Classifier," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 2, pp. 1532–1542, 2022, doi: 10.35957/jatisi.v9i2.1483.
- [11] T. N. Halim, R. Martin, and ..., "Klasifikasi Kepuasan Pelanggan Terhadap Platform E-Commerce dengan Metode K-Nearest Neighbor (K-NN)," *Jurasik (Jurnal Ris. ...)*, vol. 8, pp. 512–523, 2023, [Online]. Available: <http://ejurnal.tunasbangsa.ac.id/index.php/jurasik/article/view/636%0Ahttps://ejurnal.tunasbangsa.ac.i d/index.php/jurasik/article/download/636/609>
- [12] A. Hidayat, Mustakim, M. Z. Fauzi, and I. Syukra, "Implementasi Algoritma K-Nearest Neighbor dan Probabilistic Neural Network untuk Analisis Opini Masyarakat Terhadap Toko Online di Indonesia," *Semin. Nas. Teknol. Informasi, Komun. dan Ind. 11*, no. November, pp. 254–261, 2019.
- [13] R. Syahri and D. Puspita, "Klasifikasi Pengunjung Wisata Di Kota Pagar Alam Dengan Menggunakan Algoritma K-Nearest Neighbor (K-Nn)," *J. Ilm. Teknosains*, vol. 9, no. 2, 2023.

- [14] C. S. D. Prasetya, "Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 194, 2017, doi: 10.25126/jtiik.201743392.
- [15] T. M. S. Mulyana, "Implementasi Algoritma K-Nearest Neighbor Untuk Rekomendasi Smartphone," *Petir*, vol. 16, no. 1, pp. 29–38, 2023, doi: 10.33322/petir.v16i1.1707.